SEIR Model to assess the covid-19 pandemic behavior in Marabá-PA

Modelo SEIR para avaliação do comportamento da pandemia de Covid-19 em Marabá-PA

Walisson Ferreira Barbosa¹, Ester Barros da Costa Moreira², Juliana Mattei de Araújo³, Antônio Pazin-Filho⁴, Cláudia Dizioli Franco Bueno⁵


ABSTRACT: Objective: To determine the behavior of the COVID-19 case curve in Marabá. Methodology: The SEIR compartmentalized model was applied based on the local epidemiological data with the estimated values of latency time and infectious time obtained in Chinese populations that were tested objectively to estimate the development of SARS-CoV-2 infection in Marabá. Results: The first peak showed a total of 1438 infected (28/09/2020) after the documentation of the first case (23/03/2020) demonstrating exponential behavior. We also observed for the next 30 days, from 14/08/2020, a downward trend in the number of cases in the city, the number of basal reproduction ($R_0$) assumed the value of 3.29 between the beginning of the pandemic and the date of the first peak (28/06/2020), at the end of the study period this number was 0.8. Discussion: The model performed well when compared to the documented cases, but there was a discrepancy in the downward phase of the pandemic, possibly due to the absence of more accurate data. This adequate behavior of the model based on these data indicates that the COVID-19 dissemination should be similar despite the geographical distances and the miscigenation of the Brazilian population. Conclusion: The use of the SEIR model based on local epidemiological data with the estimated values of latency time and infectious time obtained in objectively tested Chinese populations proved to be a useful tool for predicting the behavior of the COVID-19 pandemic.

Keywords: COVID-19; Coronavirus disease; Forecasting; Statistical models; Epidemiology; Brazil/Epidemiology.

RESUMO: Objetivo: Determinar o comportamento da curva de casos de COVID-19 em Marabá. Metodologia: Aplicou-se o modelo compartimentado SEIR baseando-se nos dados epidemiológicos locais com os valores estimados de tempo de latência e tempo infeccioso obtido em populações chinesas que foram testadas objetivamente para estimar o desenvolvimento da infecção pelo SARS-CoV-2 em Marabá. Resultados: O primeiro pico apresentou um total de 1438 infectados (28/09/2020) após a documentação do primeiro caso (23/03/2020) demonstrando comportamento exponencial. Observou-se ainda para os próximos 30 dias, a partir do dia 14/08/2020, uma tendência de queda do número de casos no município, o número de reprodução basal ($R_0$) assumiu o valor de 3,29 entre o início da pandemia e a data do primeiro pico (28/06/2020), já ao final do período de estudo este número foi de 0,8. Discussão: O modelo teve bom comportamento ao se comparar com os casos documentados, mas observou-se uma discrepância na fase descendente da pandemia, possivelmente decorrente da ausência de dados mais acurados. Este comportamento adequado do modelo com base nestes dados indica que o comportamento de disseminação da COVID-19 pode ser assumido como sendo similar apesar das distâncias geográficas e da miscigenação da população brasileira. Isto abre a possibilidade de se utilizar o modelo SEIR para previsão da pandemia em territórios com carência de testagem viral. Conclusão: A utilização do modelo SEIR baseando-se nos dados epidemiológicos locais com os valores estimados de tempo de latência e tempo infeccioso obtido em populações chinesas testadas objetivamente demonstrou-se como uma ferramenta útil para previsão do comportamento da pandemia de COVID-19.

Palavras-chave: COVID-19; Coronavírus; Infecções por coronavírus; Epidemiologia analítica; Modelos epidemiológicos; Brasil/epidemiologia.

¹Universidade do Estado do Pará, Faculdade de Medicina, Departamento de Ciências Biológicas e da Saúde. ORCID: https://orcid.org/0000-0003-1450-6363. E-mail: walisson.barbosa@aluno.uepa.br.
²Universidade do Estado do Pará, Faculdade de Medicina, Departamento de Ciências Biológicas e da Saúde. ORCID: https://orcid.org/0000-0002-8163-3832. E-mail: ester.moreira@aluno.uepa.br.
³Universidade do Estado do Pará, Faculdade de Medicina, Departamento de Ciências Biológicas e da Saúde. ORCID: https://orcid.org/0000-0002-1195-4112. E-mail: juliana.araujo@aluno.uepa.br.
⁴Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto. ORCID: https://orcid.org/0000-0001-5242-329X. E-mail: apazin@fmrp.usp.br.
⁵Universidade do Estado do Pará, Faculdade de Medicina, Departamento de Ciências Biológicas e da Saúde. ORCID: https://orcid.org/0000-0001-7308-117X. E-mail: claudia.bueno@uepa.br.

Endereço para correspondência: Walisson Ferreira Barbosa. Av. Itacaiúnas, 1607. Novo Horizonte. Marabá-PA, Brasil. E-mail: walisson.barbosa@aluno.uepa.br.
INTRODUCTION

SARS-CoV-2 coronavirus is the cause of COVID-19, a disease with respiratory manifestations ranging from asymptomatic to respiratory failure. The treatment is limited to supportive measures and vaccination in the municipality of Marabá began on January 19, 2021. It has rapidly spread through respiratory or contact, with exponential growth in the number of cases. About 40 days after the first registration in Wuhan (Hubei Province, China), the World Health Organization (WHO) declared COVID-19 as a Public Health Emergency of International Importance (PHE-II), and by June 2021 confirmed more than 183 million cases and approximately 4 million deaths worldwide. The first documented case of the disease in the national territory was on February 26, 2020.

In Brazil, the assistance provided by the health system is more precarious in the Northern region, which can suffer more significant overload in the health system due to the lack of structure and professionals. Pará state has the most considerable number of COVID-19 cases in this region, with the first case reported on March 18, 2020. The city of Marabá is the 4th largest in Pará, and the headquarters of the Carajás Health Region - 11th regional health, with a population of 279,349 inhabitants and its first case reported on March 23, 2020. The Regional Hospital of Marabá became a reference for severe cases, and the City Hospital of Marabá as a triage center to fight the pandemic.

Despite the efforts to increase the capacity to face the pandemic, they were not sufficient to contain the growth of cases. Estimating this growth can be fundamental for adopting proactive measures that reduce the impacts caused by the dissemination of COVID-19. One of the available alternatives is compartmentalized mathematical models, which enable simulations of the evolution of the disease over time. Among them, there is the SEIR epidemiological model, derived from the compartmented model SIR, which we used in the present study to determine the behavior of the growth curve of COVID-19 cases in the city of Marabá-PA.

METHODS

We performed an analytical, cross-sectional observational epidemiological study, retrospective in data collection and prospective in estimating the behavior of cases, approved by the Research Ethics Committee of the State University of Pará Campus VIII / Marabá-PA (CAAE 33465720.3.0000.8607). We enrolled all individuals residing in the municipality of Marabá in the study from July 18 to August 14, 2020, based on the daily epidemiological bulletins of the City of Marabá from March 20, 2020, to August 14, 2020, and on the information available on the Civil Registry Transparency Portal.

The classical methodology of the SEIR Compartmental Model was used, which classifies the population (N) into susceptible (S), exposed (E), infected (I), removed (R). These differ daily so that St, Et, It, Rt vary depending on the variables (number of reproduction at time t), (average period of incubation), (average period of infection), i.e., as shown in Figure 1.

SEIR Compartmental Model

Based on this assumption, the SEIR compartmentalized model takes into account two distinct periods of a disease, the Latency Time (α⁻¹) and the Infection Time (γ⁻¹), for the classification of the individual in the exposed (E) and infected (I) compartments, respectively. It comprises the time interval between exposure to the virus and the beginning of transmission, while it corresponds to the period in which transmission can occur. At the same time, the transition of these infected to the removed (R) classification depends on recovery and lethality rates.

The probability in each period of time t of an individual moving to the next compartment in each period is defined successively in the equations:

\[
\begin{align*}
    p_{SE}(t) &= 1 - \exp\left(-\frac{\alpha t}{N}\right) \quad p_{EI}(t) = 1 - \exp(-\alpha) \quad p_{IR}(t) = 1 - \exp(-\gamma) \\
    p_{IE}(t) &= 1 - \exp(-\gamma) \\
    p_{EI}(t) &= 1 - \exp(-\alpha) \\
    p_{IR}(t) &= 1 - \exp(-\gamma)
\end{align*}
\]

Where \(p_{SE}(t)\) represents the probability of moving from compartment S to E, from E to I and from I to R. \(\gamma\) corresponds to the number of medium secondary infected caused by one infected in a period \(t\); \(\gamma\) symbolizes the inverse time of infection; \(\alpha\) configures as the inverse of latency time; \(\gamma\) constitutes the number of infected in a period \(t\) and \(N\) portrays the total susceptible population, in this case, the entire population of Marabá.
The $\alpha^{-1}$ variable, based on WHO and Guan et al., ranges from 4.1 and 7 days with a 95% confidence interval\(^{16,17}\). In turn, $\gamma^{-1}$ range from 10 to 14 days with a 95% confidence interval according to Chen et al.\(^{19}\). Therefore, we used as values of $\alpha^{-1} = 5.4$ days and $\gamma^{-1} = 12$ days in our model.

We used binomial distribution to determine possible values of fluxes between compartments

$$SE_{t+1} \sim \text{Binomial}(\tau_t, \gamma), \quad EI_{t+1} \sim \text{Binomial}(\alpha) \quad \text{and} \quad IR_{t+1} \sim \text{Binomial}(\gamma).$$

Thus, the compartments evolve daily according to the equations

$$S_{t+1} = S_t - SE_{t+1}, \quad E_t^{t+1} = E_t + SE_{t+1} - EI_{t+1}, \quad I_t^{t+1} = I_t + EI_{t+1} - IR_{t+1}, \quad \text{and} \quad R_{t+1} = R_t + IR_{t+1}.$$}

The reproduction number ($r_t$) is defined as the number of secondary cases generated from this first case throughout its infectious period in a range of time $t$. For each 12-day period, a period equal to the selected infection time, we assumed a different value of $r_t$ in order to circumvent the considerable inconsistency in the number of infected individuals accumulated during this period of time.

Thus, $r_t$ assumes the value that used in the differential equations of the SEIR model minimizes the $D$ value of in the objective function

$$D = \sum_{t=1}^{T} \left( \frac{C(t) - C(t)}{\sqrt{\sigma(t)}} \right)^2.$$ 

In which, $C(t)$ represents the number of infected people accumulated at time $t$ estimated by the model, that is, $C(t) = L_t + R_t$. In turn, $C_0(t)$ depicts the accumulated number of infected people notified by the city hall and $D$ numerically represents how far the values of $C(t)$ and $C_0(t)$ are out of line.

We keep on calculating the SEIR model for another 30 days to predict the behavior of the infected curve, assuming a constant $r_t$ value equal to the $r_t$ value calculated for the last 12-day period.

Furthermore, we applied a moving average of 12 days, based on the infectious period, to mitigate the fluctuations in daily data reported by the city.

We compared the behavior of our estimated infected curve according to the measures to combat the coronavirus in Marabá. The period was determined taking into account the incubation period\(^{18}\) and the infectious period as four and twelve days\(^{19}\), respectively, so the previous four days are irrelevant for changing the curve since the individuals are not yet in the infectious window.

Finally, we calculated a value of $r_0$ as a constant $r_t$ from the period from the start of the pandemic to the date of the first peak, to assess in general how fast the number of cases in Marabá grows.

**RESULTS**

The model approached the data reported by the city hall as shown in Figure 2; however, in the period indicated by the arrow, the data are discrepant. Figure 3 shows the curve of infected people and two peaks over the period studied. The first peak, dated 06/28/2020, has 1438 infected, while the second peak, dated 7/23/2020, has 912 infected.

When observing the daily notification of cases adjusted by the moving average (12 days), we observed the formation of two peaks that coincide with the first peak estimated by the model (Figure 3). When comparing the second peak estimated by the SEIR model with reported daily cases, we observed no corresponding peaks. Finally, the model shows a downward trend in the number of infected people over the next 30 days. The $r_0$ assumed the value of 3.29 between the beginning of the pandemic and the date of the first peak (06/28/2020), already in the last weeks this value has fallen to 0.8.

**Figure 2.** The figure shows the comparison of the accumulated infected estimated by the SEIR model with the accumulated infected notified by the city government and a forecast for the next 30 days of the accumulated infected, both with a 95% confidence interval.
The arrow indicates a darker region in which the model is more discrepant, where, despite the downward trend in the number of cases observed in the Marabá, the calculations performed by the model show an upward trend in the same period.

**Figure 3.** The figure shows the comparison of the number of infected cases estimated by the SEIR model (left axis) with daily-infected cases documented by the Municipality of Marabá (right axis).

In the first axis (left), a curve of infected by SARS-CoV-2 in Marabá and a forecast for the next 30 days estimated using the SEIR model with a 95% confidence interval. On the second axis (right), the number of daily-cases reported by the city and its moving average over a 12-day period. The arrow indicates a darker region where a second peak occurs, which can be considered a mathematical anomaly and limitation of the model used.

**DISCUSSION**

We used the SEIR model to analyze the behavior of COVID-19 in a city in the northern region of Brazil with few resources for testing and coping with the pandemic. We constructed the model based on data provided by the health authorities. The model behaved well compared to the documented cases, but there was a discrepancy in the descending phase of the pandemic, which may perhaps be due to the lack of more accurate data.

The cumulative cases estimated by the SEIR model showed a good correlation with data from health authorities, considering that we based the calculations on these data. Still, it is essential to emphasize that we used the estimates for the measures based on data from two Chinese studies with many patients and which performed testing that was not possible in Marabá\(^{17,18}\). The acceptable behavior of the model is crucial as it indicates that COVID-19 dissemination can be assumed to be similar despite geographic distances and miscegenation, a demographic trait of the Brazilian population, even though China has an ethnic race predominantly homogeneous. If it is true, it opens up the possibility of using the SEIR model to predict the pandemic in territories with a lack of viral testing, as is the reality in most Brazilian cities.

Figure 3 shows, in Marabá, the peaks of COVID-19 cases, representing the moment when the number of new cases tends to decrease or stabilize. It is worth noting that the projections of the chain of contagion by the disease can constantly vary over time, considering the new conditions of the epidemic and adherence to measures to combat the spread of the virus. It is noteworthy that after the epidemiological peak if there is a reduction in the number of cases, it does not mean that the curve will remain to descend in terms of the number of infections. Still, on the contrary, a new phase of virus propagation may occur, resulting in the return of ascendancy of the curve\(^{20}\). We will have adjust calculations for new conditions such as new waves of contagion facilitated by the absence of vaccines, validated pharmacological treatments, and flexible measures and other strands emerge.

Among the actions that contributed to the epidemic peak, there is the influence of decree number 60, of June
18, 2020, which deals with the permission to reopen some services such as shopping malls, gyms, private higher education entities, vocational courses, and religious activities such as masses and cults. We observed that partial reopening contributes to the increase in the number of cases. Decree number 61, dated June 18, 2020, also establishes specific rules for the opening of trade that must follow particular hours of operation and maintain health surveillance rules. In addition, decree number 66 deals with capacity restrictions for gyms, restaurants, and their counterparts and the reinforcement of distance and other health regulations.

As indicated by the arrow in Figure 2, despite the downward trend in the number of cases observed in the city, the calculations performed by the model show an upward trend in the same period. As a result, we follow this second peak in the period of the most significant discrepancy between the actual and estimated cases, indicated by the arrow in Figure 3. Based on the available data, it is difficult to explain this discrepancy objectively. One possibility could be the reduction in the documentation of cases by the health authorities, considering that the population with milder cases has not sought health services because they have already learned on the behavior of the pandemic. However, another possibility could be that this “new” peak is a distortion and, therefore, a possible limitation for using the model. However, considering the excellent performance observed, we will need new data to explain this distortion and discourage using the model with all the potential of a proactive tool for estimating cases.

According to Renardyet al.\textsuperscript{22}, the reduction of individual contacts can delay the occurrence of the second peak and reduce its magnitude, which may be happening in Marabá considering that the estimated number of infected people is close to that estimated by these authors.

The value of $r_f$ assumed for the next 30 days as of 08/14/2020 was equal to the calculated for the last 12 days, that is, equal to 0.8. A value that is below the threshold of 1 stipulated in the literature with a minimum value to maintain epidemic growth\textsuperscript{23} indicating, as shown in Figure 3, a downward trend.

To achieve herd immunity in places with a value $r_f$ greater than 2, the immunized population must represent 60-80%, this population fraction being directly proportional to the value of $r_f$\textsuperscript{24}. However, Aguas et al. noted that when the basic reproduction number ($R_0$) assumes values between 2.5 and 5, the percentage of individuals needed to reverse the curve and avoid new peaks drops to 10-20%\textsuperscript{25}.

In Marabá, the figure $R_0$ found between the first notified case and the peak is approximately 3.29. The $R_0$ finding for the municipality is 0.03 point higher than that found for Brazil in Kwok et al.\textsuperscript{26}, who inferred that at least 69.3% of the infected population would be needed for the emergence of a new wave to be avoided or mitigated. However, in the final period of analysis, a decreasing trend is observed, which may indicate that this percentage may be lower than that previously inferred by Aguas et al.\textsuperscript{25}.

Thus, the present study provides further understanding the dynamics of SARS-CoV-2 transmission in the city of Marabá-PA, which could be helpful in the planning of interventions by the epidemiological surveillance team, and for the promotion of improved assistance to infected, in addition to serving as a tool for containing the virus. The use of the SEIR model facilitates comparison with other studies as it is a well-known model in the literature. In addition, it provides a basis for future studies in the region on the subject, such as the analysis of underreporting by other researchers.

The study’s limitations refer to the unavailability of recent data on the number of inhabitants of Marabá, so we used the most current period available – 2018 - in the model. We used variables (incubation time and infectious period) based on studies carried out in China, so one should consider climatic differences, seasonality, and viral mutations when comparing studies. In addition, official daily bulletins published by the City Hall and the municipal health authorities, considering that the population with milder cases has not sought health services because they have already learned on the behavior of the pandemic. However, another possibility could be that this “new” peak is a distortion and, therefore, a possible limitation for using the model. However, considering the excellent performance observed, we will need new data to explain this distortion and discourage using the model with all the potential of a proactive tool for estimating cases.

According to Renardyet al.\textsuperscript{22}, the reduction of individual contacts can delay the occurrence of the second peak and reduce its magnitude, which may be happening in Marabá considering that the estimated number of infected people is close to that estimated by these authors.

The value of $r_f$ assumed for the next 30 days as of 08/14/2020 was equal to the calculated for the last 12 days, that is, equal to 0.8. A value that is below the threshold of 1 stipulated in the literature with a minimum value to maintain epidemic growth\textsuperscript{23} indicating, as shown in Figure 3, a downward trend.

To achieve herd immunity in places with a value $r_f$ greater than 2, the immunized population must represent 60-80%, this population fraction being directly proportional to the value of $r_f$\textsuperscript{24}. However, Aguas et al. noted that when the basic reproduction number ($R_0$) assumes values between 2.5 and 5, the percentage of individuals needed to reverse the curve and avoid new peaks drops to 10-20%\textsuperscript{25}.

In Marabá, the figure $R_0$ found between the first notified case and the peak is approximately 3.29. The $R_0$ finding for the municipality is 0.03 point higher than that found for Brazil in Kwok et al.\textsuperscript{26}, who inferred that at least 69.3% of the infected population would be needed for the emergence of a new wave to be avoided or mitigated. However, in the final period of analysis, a decreasing trend is observed, which may indicate that this percentage may be lower than that previously inferred by Aguas et al.\textsuperscript{25}.

Thus, the present study provides further understanding the dynamics of SARS-CoV-2 transmission in the city of Marabá-PA, which could be helpful in the planning of interventions by the epidemiological surveillance team, and for the promotion of improved assistance to infected, in addition to serving as a tool for containing the virus. The use of the SEIR model facilitates comparison with other studies as it is a well-known model in the literature. In addition, it provides a basis for future studies in the region on the subject, such as the analysis of underreporting by other researchers.

The study’s limitations refer to the unavailability of recent data on the number of inhabitants of Marabá, so we used the most current period available – 2018 - in the model. We used variables (incubation time and infectious period) based on studies carried out in China, so one should consider climatic differences, seasonality, and viral mutations when comparing studies. In addition, official daily bulletins published by the City Hall are a secondary data source and can be due to considerable bias, such as underreporting.

CONCLUSION

In short, using SEIR model, based on local epidemiological data with estimated values of latency time and infectious time obtained in external objectively tested populations, proved a reliable tool for predicting the behavior of the COVID-19 pandemic. These models could be of use in places lacking capability of testing, as is the reality in most Brazilian municipalities.
REFERENCES


Submeted: 2020, November 04
Accepted: 2021, July 30